

The Effect of Pose on the Distribution of Edge Gradients in Omnidirectional Images

Dean Jarvis¹ and Theodoros Kyriacou²

¹d.g.jarvis@keele.ac.uk

²t.kyriacou@keele.ac.uk (Corresponding/Senior Author)

School of Computing and Mathematics, Keele University, Staffordshire, ST5 5BG, UK

Abstract. Images from omnidirectional cameras are used frequently in applications involving artificial intelligence and robotics as a source of rich information about the surroundings. A useful feature that can be extracted from these images is the distribution of gradients of the edges in the scene. This distribution is affected by the pose of the camera on-board a robot at any given location in the environment. This paper investigates the effect of the pose on this distribution. The gradients in the images are extracted and arranged into a histogram which is then compared to the histograms of other images using a chi-squared test. It is found that any differences in the distribution are not specific to either the position or orientation and that there is a significant difference in the distributions of two separate locations. This can aid in the localisation of robots when navigating.

1 Introduction

The goal of this paper is to test a bio-inspired system that can recognise and distinguish between two locations¹ using visual information, specifically information from edges within the scene. Tests will measure how changes in pose may effect this information and whether or not there is a significant difference between separate locations. Tests are made simpler by using an omni directional camera. Visual information is used as it is the primary sense that mammals use to navigate in their environments. The omnidirectional camera emulates the near omnidirectional vision of rodents and to reduce the amount the variability of images that can be taken from same position.

In the field of robotics there are many parts of a design one might want to consider. One very important ability that is required for an autonomous robot would be to navigate within its environment. Navigation can be broken down

¹ During the course of this paper the word "location" will refer to the immediate environment (the set of positions that share the same contextual name) of the robot for example the kitchen or the lab. Whereas "position" will be an exact measure of where the robot is (eg coordinates x, y). Pose defines the position (x, y) and orientation (theta) of the robot.

into three distinct tasks, localisation, mapping, and path finding [1]. There are in essence two ways to approach robot navigation; there are the *pure* computational methods and biologically inspired methods. Pure methods include examples such as those described in [2],[3], and [4]. These methods have come a long way in producing reliable results for navigation but they often produce a lot of data and high fidelity maps that then require optimisation before they can be used efficiently(see [3]). This means that these methods have a high demand for memory and computational power. In contrast animals seem to do the same tasks without high fidelity maps, using primarily visual stimuli. Biologically inspired robotic navigation can be further broken down into two more categories; top down methods and bottom up methods. For example top down methods start by observing the behaviour of living organisms and how they react to stimuli and situations, whilst attempting to reproduce these behaviours in robots. Franz and Mallot demonstrate a review of some top down biologically inspired robot navigation methods, which take inspiration from insects and rodents [5]. In contrast bottom up methods attempt to reproduce natural navigation by looking at how individual parts of an animal may work in tandem to navigate in an environment. This could be how individual neurones fire in response to stimuli within their receptive field. Work produced by Kyriacou (see [6]) is one such example; this work focuses on producing a working computational model of head direction cells within the brain that aid in determining the orientation of the robot.

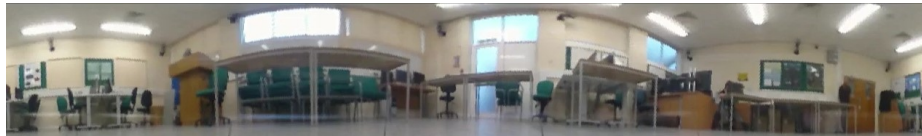
Bio-inspired computing has influenced the way in which robots localise and navigate in their environments. Sensors designed to mimic the way insects use polarised light from the sky to navigate to and from their nests as described in work by [10]. Other notable examples include teaching a system to change navigation methods based on environmental cues and the concept of Time Till Contact (TTC) which can be utilised for obstacle avoidance without feature detection by using optical flow. This can be improved by incorporating angular velocities and a Kalman filter [11, 12, 13]. Localisation using vision is essentially the ability to compare the current view to images in memory that are linked to a spatial location [14]. It has been found and discussed by Anzai [15] how the visual system in monkeys exhibits selective firing when encountered edges and textures that correspond to certain angles that exist within the visual field within the V2 region of the brain. The cells within V2 region of the visual system respond more to complex stimuli such as curves and angles. This suggests that this information is an important component of the image, the image itself being important for localisation. By developing a localisation system based on information gathered from the behaviour of the brains of animals it is hoped that the resulting system will be robust.

There have been studies that use edges and gradient information from images to compute image similarity [16][17] and localise [18]. However there is a lack of information about how this kind of information changes with respect to the pose of the camera and whether or not it can be used as a stand-alone method to differentiate between locations in the navigable environment.

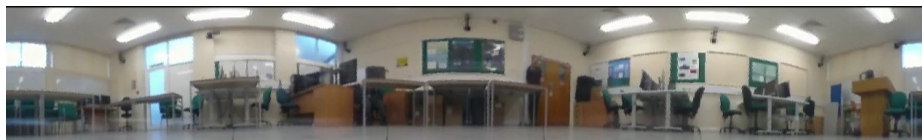
Omnidirectional images could be used to compare the views taken at different locations. Images taken using omni-directional cameras are unwrapped in such a way so as to form a continuous panorama image. During the unwrapping process there is some warping of the original image, distorting the information density, due to the nature of the omni-directional camera lens and the interpolation methods that have to be employed [19]. Omni directional images used in applications that also use image gradient distribution could potentially suffer with inaccurate comparisons due to changes in camera pose. Work in object recognition with omni-directional cameras, for example [20], could be affected by these changes in pose. There is little available research on the effect that a change in pose can have on the gradient distribution of an omnidirectional image. This study explores the effect of changing pose on the gradient distribution by comparing similarity of the histograms of gradients in images using the chi-squared test [22] as a metric of similarity.



(a) Reference position image for empirical analysis.



(b) Image after moving camera forward 80cm showing an apparent distortion on the tables.



(c) Image after rotation the reference image by 60 degrees.

Fig. 1: These photographs show the unwrapped omnidirectional images taken at two different positions in the same room. At first glance it seems that there is a greater apparent change in the images between the change in position compared to the change in orientation

It is hypothesised by looking at images such as those in Figure 1, that there will be a greater difference in the gradient histograms due to position than there will be due to the orientation as the camera lens is rotationally symmetric and a change in position seems to cause a greater distortion. Also there will be a

significant difference between the histograms of gradients between two different locations (between Figs. 1&4) that could be utilised as a determinant of location.

The aim of this study is to see how, by using the gradients of edges, a scene may be used to localise a robot within it's environment using only visual information received from it's camera. In order to test this, omni-directional images taken in the same location (room) but from different poses (in terms of position and orientation) are compared with each other by means of their edge content. Also, images taken in another location (i.e. a different room) are compared with the first set using the same method.

2 Method



Fig. 2: This image shows the camera with 360 degree lens attachment mounted with velcro on the back of an RC car.

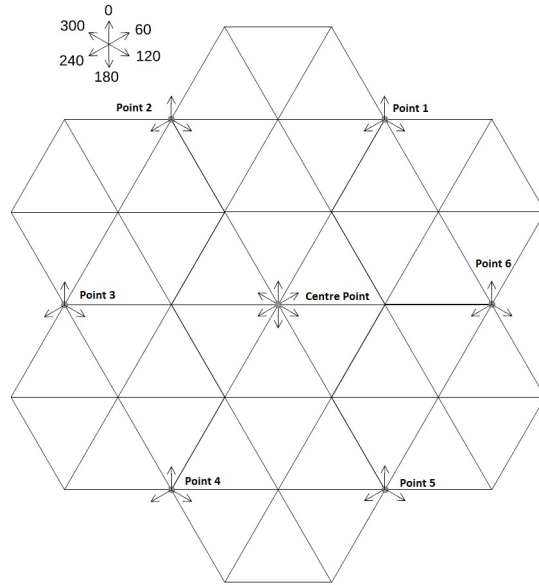
A Sony Bloggie ² camera with a 360 lens attachment, as shown in Figure 2, was used to capture all the images that were used in this study. A limited perspective camera was not used due to the fact that locations would require the capture of multiple views from which to be identified. Many animals benefit from wild fields of views that allow them to be aware of more of their environment without having to move around a lot to do so. An omnidirectional camera can handle this issue better as all possible views of a position are incorporated into a single image. The room where a majority of the images were captured (GR) conveniently had a floor of tiled equilateral triangles as shown in Figures 3a&3b. This room was chosen for this reason to ensure that distances and angles between images were constant. Initially, to check if there was any inherent time dependant error due to the hardware, software, or the environment, multiple images were taken from the same pose (Figure 3b centre point at 0 degrees). Images taken from this one

² <https://www.sony.co.uk/electronics/support/webbie-hd-bloggie-cameras-mhspm-series/mhs-pm5> Last accessed: 1/05/2018

pose were used to compute the minimum difference that could be attributed to temporal error. The Positions that images were taken from are also described in Figure 3b, which shows the 7 positions and the various angles at which the images were taken for a total of 24 images for the GR. The centre point has smaller angle intervals as this is used primary to test for changes due to orientation without changes in position. Whereas points 1-6 are used to test for changes due to position, and changes due to changes in position and orientation.



(a)



(b)

Fig. 3: (a) Shows the GR and its convenient tiled flooring. (b) Shows a partial layout of the first room images were taken with all the positions and orientations marked. Length of triangle sides are 40cm.

For future comparison, another set of images were taken from a different room (TR), see Figure 4 using the same angle increments as the centre point in the GR. The angular increments were kept constant with the use of a regular hexagon template. This is also another point of comparison for how changes in orientation alter the edge gradient distribution.

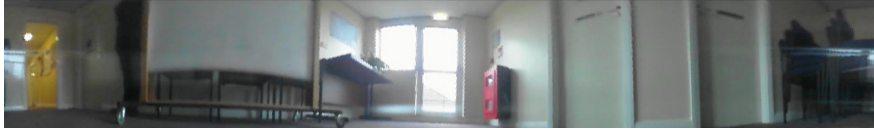
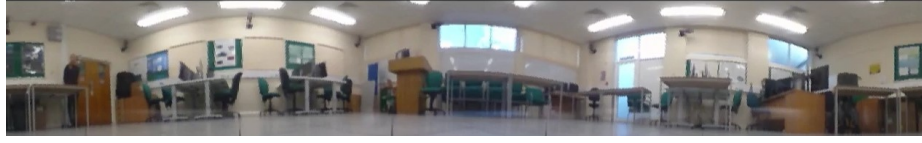


Fig. 4: This is an image of the TR from which images were used to distinguish the viability of this method as one that would be suitable to distinguish between different locations.

2.1 Image Processing

To process the images the OpenCV ³ library was used in conjunction with the Sony Playmemory Home software. Playmemory Home was used to unwrap the raw images into a panorama like image. OpenCV was then used to extract the gradient information from the image. To do this, Sobel operators [21] in the x and y direction were used to get two gradient images, and after this a method was used to combine these images to produce two new images. One of the resulting images being the image where each pixel value represents an angular value between 0 and 360 which is also the direction of the edge as shown in Figure 5c. The other image being a magnitude image where each pixel value is the magnitude of the gradient as shown in Figure 5b. The magnitude image was altered via a binary threshold where the cut off point was one fifth the maximum magnitude. This threshold was empirically chosen. The chosen threshold appeared to be a good value where prominent edges persist whilst artefacts due to low magnitude edges such as textures were removed. The resultant binary image was used as a mask to remove noisy gradients with low magnitudes that do not provide any useful information in this context. Once the gradient image had been masked the value of every non zero pixel was written to a file for comparison against other images.

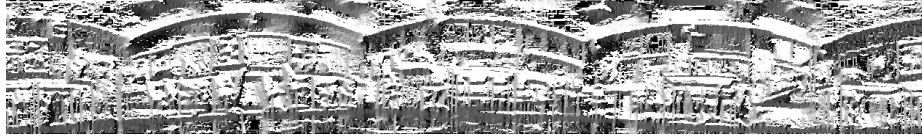
³ <https://opencv.org/> Last accessed: 1/05/2018



(a) Unwrapped image before processing.



(b) Full magnitude image.



(c) Full angle image.



(d) Binary threshold of magnitude image.



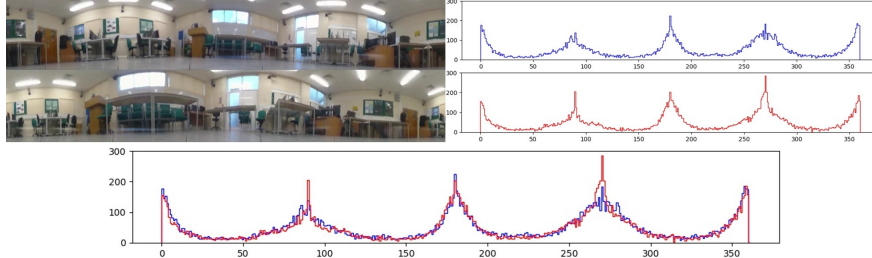
(e) Angle image after being masked with the binary threshold. This shows the location and direction of edges in degrees.

Fig. 5: This figure shows intermediate image Processing Steps. Images (b) and (c) are obtained from (a). Image (d) is obtained via a binary threshold applied to image (b). Image (d) is used to select important information from image (c). Image (e) is the final product used to produce the gradient distribution

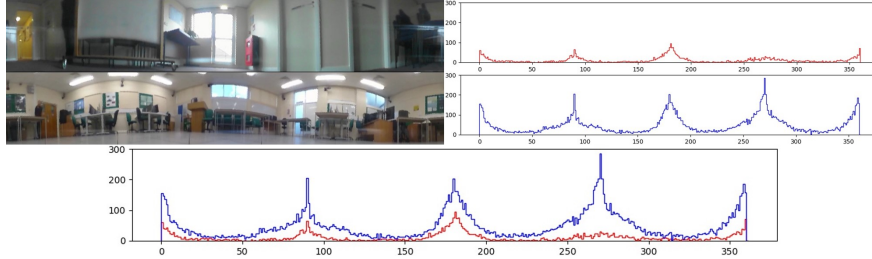
2.2 Data Representation

The edge information from the collected images was used to produce histograms that describe the gradient distribution at each of the sampled positions in the two locations used. These histograms were then compared using openCV's *compHist* method. The *compHist* method has four different tests it can use to compare the similarity of the histograms. The one that was chosen to compare the histograms was the chi-squared test [22] where big values of the test statistic (chi-square) mean big differences (in terms of edge distribution) in any two images being

compared. Figs. 6a and 6b illustrate how the chi-squared value represents a distance between the histograms of seemingly similar and different places.



(a) This Figure shows how two seemingly similar places look when their histograms are overlaid. This would give a very small chi-squared metric as there is a small difference between the lines at any point.



(b) This Figure shows how two seemingly different places look when their histograms are overlaid. This would give a large chi-squared metric as there is a large difference between the lines at many points.

Fig. 6: The x axis of the histograms are values that gradients can take in the image. The y axis is how many time a given gradient occurs.

3 Results

The check for temporal errors resulted in some low level differences across different frames taken from the same pose. The mean chi-squared distance from images at the same position and orientation was 860.8 ± 98.8 , this information gives an estimate of how much error there may be within any other results due purely to external factors, e.g: camera auto exposure calibration(no camera option to disable), light changes due to flickering lights, and sensor noise on the camera.

The main aim of this work was to check whether changes in pose are more dependant on position or orientation, therefore a comparison of the image data where only the position and only the orientation was changed was performed. To do this, for every image recoded the histogram generated from it was compared to

that of the other images, resulting in a table of chi-squared results that could be used to easily visualise any relationships between the different poses from which the images were captured. The table in Figure 7 is illustrated visually (using grey-scale to represent the chi-square value) . A t-test was performed to check whether there was a significant difference between images where only position was changed and images where only orientation was changed, this was done using images from both the GR and the TR. To do this all the data points that are the result of positional changes only were averaged and data points that are the result of changes in orientation only were also averaged. These averages were used to perform the T-test. The average chi-squared distance due to changes in position was 1947 ± 1448 whereas the average chi-squared distance due to changes in orientation only was 1851 ± 1005 . A rejection of the leading hypothesis is attained showing that there is no significant difference between the means ($p = 0.393 > 0.05$).

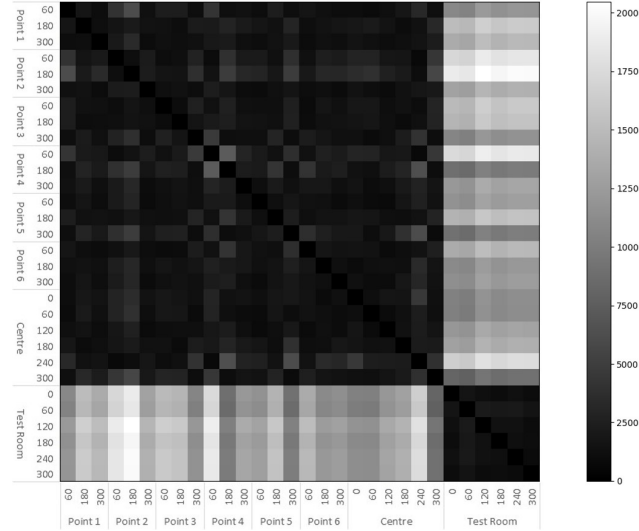


Fig. 7: This image shows how every image was compared to all other image. Each pixel represents a chi-squared value, resulting from a comparison of two rooms which can be found on the axis. It highlights the distinction between similar and dissimilar rooms where low values-dark correspond to comparisons of images captured at different poses in the same room (GR) and high values-light correspond to comparisons of images taken in different rooms (GR&TR). The axes of this illustration refer to the position and orientation of the corresponding image; e.g Point 1 60 is the image taken from Point 1 at 60 degrees as shown in Figure 3b.

Using the same method of comparison (i.e the t-test), all the data points from the Figure 7 that corresponded to comparisons of images taken in different rooms (i.e the light regions) were averaged and compared to the average of the data points from image comparisons of the same room (i.e the dark regions). The mean difference between images from the same room was 1917.8 ± 1062 and the mean from images of different rooms was 13574.1 ± 2920 . Due to this data the alternate hypothesis, that there will be a significant difference between the histograms of differing rooms, fails to be rejected as there is a significant difference between the means of the two sets of data ($p = 2.6 \times 10^{-95} < 0.05$).

Conclusion and Further Work

It was found that a change of position seemed to have a similar difference to a change in orientation; this was shown within the results section with the t-test showing no significant difference between the data from each. It has also been confirmed how gradient histograms as a global feature could potentially be used as a distinguishing feature between two locations. This global feature could be used as a coarse localisation feature as a fast way to limit any search space for finer localisation later. However, more research would be required to further justify this statement. This could be done by comparing images taken from multiple locations in the environment of the robot in the same way as this work has shown. How will a larger set of areas affect the robots ability differentiate between locations? The method discussed on the paper has not been designed to handle transitional locations. For example, if the robot were in a door way and could see 2 separate locations where would it say it belongs? Future work could look at developing a strategy to handle this scenario. This could possibly be accomplished by segmenting the image if there are two locations with similar chi-squared distances. Future work could also use video as opposed to still images will be used, providing multiple images per position but also the temporal link between them. This could mitigate any false positives due to a poor comparisons as the incorrect comparison would be overshadowed by the average from the other images.

Bibliography

- [1] Fuentes-Pacheco, J., Ruiz-Ascencio, J., Rendón-Mancha, J.M.: Visual simultaneous localization and mapping: a survey. *Artificial Intelligence Review* **43**(1) (2015) 55–81
- [2] Wen, S., Zhang, Z., Ma, C., Wang, Y., Wang, H.: An extended kalman filter-simultaneous localization and mapping method with harris-scale-invariant feature transform feature recognition and laser mapping for humanoid robot navigation in unknown environment. *International Journal of Advanced Robotic Systems* **14**(6) (2017) 1729881417744747
- [3] Vallve, J., Sola, J., Andrade-Cetto, J.: Graph slam sparsification with populated topologies using factor descent optimization. *IEEE Robotics and Automation Letters* (2018)
- [4] Lowe, T., Kim, S., Cox, M.: Complementary perception for handheld slam. *IEEE Robotics and Automation Letters* **3**(2) (2018) 1104–1111
- [5] Franz, M.O., Mallot, H.A.: Biomimetic robot navigation. *Robotics and autonomous Systems* **30**(1-2) (2000) 133–153
- [6] Kyriacou, T.: An implementation of a biologically inspired model of head direction cells on a robot. In: *Conference Towards Autonomous Robotic Systems*, Springer (2011) 66–77
- [7] Laschi, C., Cianchetti, M., Mazzolai, B., Margheri, L., Follador, M., Dario, P.: Soft robot arm inspired by the octopus. *Advanced Robotics* **26**(7) (2012) 709–727
- [8] Crespi, A., Ijspeert, A.J.: Amphibot ii: An amphibious snake robot that crawls and swims using a central pattern generator. In: *Proceedings of the 9th international conference on climbing and walking robots (CLAWAR 2006)*. Number BIOROB-CONF-2006-001 (2006) 19–27
- [9] Meisel, M., Pappas, V., Zhang, L.: A taxonomy of biologically inspired research in computer networking. *Computer Networks* **54**(6) (2010) 901–916
- [10] Karman, S.B., Diah, S.Z.M., Gebeshuber, I.C.: Bio-inspired polarized skylight-based navigation sensors: A review. *Sensors* **12**(11) (2012) 14232–14261
- [11] Dollé, L., Sheynikhovich, D., Girard, B., Chavarriaga, R., Guillot, A.: Path planning versus cue responding: a bio-inspired model of switching between navigation strategies. *Biological cybernetics* **103**(4) (2010) 299–317
- [12] Horn, B.K., Fang, Y., Masaki, I.: Time to contact relative to a planar surface. In: *Intelligent Vehicles Symposium, 2007 IEEE*, IEEE (2007) 68–74
- [13] Zhang, H., Zhao, J.: Bio-inspired vision based robot control using featureless estimations of time-to-contact. *Bioinspiration & biomimetics* **12**(2) (2017) 025001

- [14] Paul, C.M., Magda, G., Abel, S.: Spatial memory: Theoretical basis and comparative review on experimental methods in rodents. *Behavioural brain research* **203**(2) (2009) 151–164
- [15] Anzai, A., Peng, X., Van Essen, D.C.: Neurons in monkey visual area v2 encode combinations of orientations. *Nature neuroscience* **10**(10) (2007) 1313
- [16] Liu, A., Lin, W., Narwaria, M.: Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing* **21**(4) (2012) 1500–1512
- [17] Hu, R., Collomosse, J.: A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *Computer Vision and Image Understanding* **117**(7) (2013) 790–806
- [18] Kröse, B.J., Booi, O., Zivkovic, Z., et al.: A geometrically constrained image similarity measure for visual mapping, localization and navigation. In: EMCR. (2007)
- [19] Jacquey, F., Comby, F., Strauss, O.: Non-additive approach for omnidirectional image gradient estimation. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, IEEE (2007) 1–6
- [20] Cinaroglu, I., Bastanlar, Y.: A direct approach for object detection with catadioptric omnidirectional cameras. *Signal, Image and Video Processing* **10**(2) (2016) 413–420
- [21] Sobel, I.: An isotropic 3×3 image gradient operator. *Machine vision for three-dimensional scenes* (1990) 376–379
- [22] Greenwood, P.E., Nikulin, M.S.: A guide to chi-squared testing. Volume 280. John Wiley & Sons (1996)